Al-Qadisiyah Journal of Pure Science
ISSN(Printed): 1997-2490 ISSN(Online): 2411-3514
DOI: 10.29350/jops
Al-Qadisiyah Journal of Pure Science

Weighted EM Algorithm for Laplace Mixture Regression

Hassan Sami Uraibi[1], Hassanin Adel Salih[2]

Received: 29/01/2026; Revised: 17/03/2026; Accepted: 17/03/2026.

Abstract

Fitting mixture regression models has received attention from the authors in the statistical literature over the last decade. That is due to its ability to model data with more than one pattern. The main reason is the latent variables, which are the source of these patterns appearing. The EM algorithm is considered the enhanced approach to the MLE estimation method for estimating its parameters. However, this method is not resistant to the outliers; heavy-tailed distributions are recommended to fit the mixture regression model. Fitting a regression model with a Laplace mixture distribution was proposed in the literature to be robust against outliers and leverage points. Whatever, the previous studies sought to trim the leverage points, while this manuscript seeks to weight the leverage points instead of trimming them. We noted that the mixture model has four types of outliers, which are considered in the simulation study. The performance of the proposed method, W.MixLap, is compared with MixLap, mixt, and MLE, which are fitted regression models with Laplace, Student, and Gaussian mixture distributions, respectively. The result shows that the performance of W.MixLap is better than others.

Keywords: Mixture regression, Outlier, Leverage point, Laplace distribution, EM algorithm

1. Introduction

Finite mixture regression models are strong statistical tools to model data heterogeneity. Quandt [13] noted that economic and social data frequently come from several normal subpopulations, with different models among variables. The traditional statistical estimation methods, such as the MLE cannot be reliable, because there may be many local maxima and flats in the likelihood surface, which makes it difficult to get the optimum solutions. Dempster, Laird, and Rubin [4] reformulated the EM algorithm as a specific solution for estimating a Gaussian mixture regression. Unfortunately, a single outlier can exhibit such a major influence on EM algorithm parameter estimates as to create spurious components [1]. McLachlan[12] suggested using a symmetric t-distribution instead of a normal mixture, and then the authors in the statistical literature

extended it to the context of regression. Lin et al. [11] pointed out that "forcing symmetric t-distributions on asymmetric data can be as bad as assuming normality." Actually, in practice, real data can display both heavy tails and asymmetry. They utilized the skewed t-distribution with density instead of.

Although a symmetric t-distribution for mixture regression of [17] far outperformed in robustness, according to Song et al [15] "parameter estimation for the degree of freedom parameters introduces a computational complexity that may not be practical when dealing with large-scale applications." Therefore, they proposed Laplace distribution mixtures, and they used the normal-exponential scale mixture representation of the Laplace distribution. The generalized systematic framework for outlier identification in mixture regression models is an academic collaboration across the various statistical subfields. Groundbreaking contributions to the area were made in regression diagnostics by [3] and [14], where it was recognized that a key step in standard linear regression is the identification of outlying data points on both response or regressor space. Yu et al. [18] who identified four types: outliers at response, outliers (leverage points) at covariates mixed of both and evidence of overlapping data points separately.

In this context, Song et al. [15] have taken trimming leverage points as a step before using data with their proposed method to get rid of the effect of leverage points. Indeed, the approach of trimming outliers in mixture model suggested by [9],[10] suggested to distinct between outliers and all other observations 'within-component outliers' and then trimmed clustering from the total components [6], [5],[7] Later, [2] developed ``robust fitting of mixture regression models" (An extension of TLE-type trimming) to deal with the case when there is contamination in data sets by trimming non-contaminated 'outliers' or leverage points (i.e., X-direction outliers) and showed a beneficial result through removing model under-performing conditions. [16] and [8] created weight in the FTLE framework. Rather than "in or out" hard trim, WFTLE provides continuous weights to data points. This weight is built from a robust measure of the residual and leverage for each point in x-space. Very large residuals (clear outliers) correspond to points with a weight close to 0. Leverage points (Values of x with extreme leverage) that do not fit the model well are downweighed; hence, it protects against bad leverage points. If a "good" point is located in the middle of the data, it is given a weight close to 1. As a robust option for dealing with the leverage points problem, The rest of this paper is organized into five sections: Section 2 presents the EM Algorithm for Laplace Mixture Regression. The Robustness of Laplace EM in Section 3, Section 4 is assigned for Weighted of EM for Laplace Mixture Regression, Simulation study, Conclusion and Recommendation are present in the Section, 5,6, and 7, respectively.

1. EM Algorithm for Laplace Mixture Regression

assume the data are modeled by various regression models, each with Laplace errors. The full model definition of Laplace Mixture regression, in likelihood of a single observation:

$$f(y_i|x_i,\theta) = \sum_{k=1}^{K} \pi_k \cdot \text{Laplace}(y_i|x_i^{\top}\beta_k, \sigma_k) \qquad (1)$$

where: π_k is mixing proportion for component k ($\sum \pi_k = 1$), β_k is regression coefficients for component k, σ_k is scale parameter for component k, and

$$\text{Laplace}(y|\mu,\sigma) = \frac{1}{2\sigma}\exp\left(-\frac{|y-\mu|}{\sigma}\right) \qquad (2)$$

This is to say that if the estimator has a higher breakdown point than its Gaussian counterpart, then it also will be more robust which follow fact that Laplace distribution attains large deviations with higher probability. The EM algorithm involves two steps: the expected step (E-step) that computes the probability an appropriate full-data measurement would be observed, and the Maximizing step (M-step), which considers a generalized form of the original problem over which to maximize. Song et al. (2014) presented a robust EM method under the Laplace error distribution, which is written as follows:

Let $z_j \in \{1,\dots,K\}$ be the latent component indicator:

$$\mathcal{L}_c(\Theta) = \prod_{j=1}^{n}\prod_{i=1}^{K}\left[\pi_i \cdot \frac{1}{2\sigma_i}\exp\left(-\frac{|y_j - x_{j'}\beta_i|}{\sigma_i}\right)\right]^{\mathbb{I}(z_j=i)} \qquad (3)$$

Initialize $\pi_i^{(0)} = 1/K$ for $i = 1,\dots,K$, $\beta_i^{(0)}$ preliminary regression, $\sigma_i^{(0)}$ or estimate from residuals. The EM algorithm can be described in two steps,

Iterate until convergence:

E-Step: For $i = 1,\dots,K$ and $j = 1,\dots,n$:

$$\tau_{ij}^{(k)} = \frac{\pi_i^{(k)} \cdot \frac{1}{2\sigma_i^{(k)}}\exp\left(-\frac{|y_j - x_{j'}\beta_i^{(k)}|}{\sigma_i^{(k)}}\right)}{\sum_{l=1}^{K}\pi_l^{(k)} \cdot \frac{1}{2\sigma_l^{(k)}}\exp\left(-\frac{|y_j - x_{j'}\beta_l^{(k)}|}{\sigma_l^{(k)}}\right)} \qquad (4)$$

In this stage the algorithm is supposed to decide whether we have a particular observation which comes from a kth regression component. It is the step of computing the posterior probabilities $\tau_{ij}^{(k)}$ from the prior ones $\pi_i^{(k-1)}$.

M-Step: this step, updating the prior information mixing proportion $\pi_i^{(k)}$, Scale parameters and regression coefficients $\beta_i^{(k)}$ as follows,

$$\pi_i^{(k+1)} = \frac{1}{n}\sum_{j=1}^{n} \tau_{ij}^{(k)}$$

$$\sigma_i^{(k+1)} = \frac{\sum_{j=1}^{n} \tau_{ij}^{(k)} |y_j - x_{j'}\beta_i^{(k)}|}{\sum_{j=1}^{n} \tau_{ij}^{(k)}}$$

The regression coefficients $\beta_i^{(k+1)}$ are obtained by solving:

$$\beta_i^{(k+1)} = \underset{\beta_i}{\operatorname{argmin}} \sum_{j=1}^{n} \tau_{ij}^{(k)} |y_j - x_{j'}\beta_i|$$

This is a weighted median regression problem can be solved via Iterative Reweighted Least Squares (IRLS) approach each inner iteration t we can solve

$$\beta_i^{(t+1)} = \left(\sum_{j=1}^{n} \tau_{ij}^{(k)} w_j^{(t)} x_j x_j'\right)^{-1} \left(\sum_{j=1}^{n} \tau_{ij}^{(k)} w_j^{(t)} x_j y_j\right) \qquad (5)$$

Where $w_j^{(t)} = \frac{\sigma_i^{(t)}}{\sqrt{2}|y_j - x_j'\beta_i^{(t)}|}$

where $\sqrt{2}$ is a scale factor to connect Laplace distribution to Gaussian approximation. Compute the log-likelihood:

$$\mathcal{L}(\Theta^{(k)}) = \sum_{j=1}^{n} \log\left(\sum_{i=1}^{K} \pi_i^{(k)} \cdot \frac{1}{2\sigma_i^{(k)}} \exp\left(-\frac{|y_j - x_{j'}\beta_i^{(k)}|}{\sigma_i^{(k)}}\right)\right)$$

Stop when $\frac{|\mathcal{L}(\Theta^{(k+1)}) - \mathcal{L}(\Theta^{(k)})|}{|\mathcal{L}(\Theta^{(k)})|+1} < \epsilon$, where $10^{-6} \leq \epsilon \leq 10^{-8}$

Though weighted median regression in the M-step is more expensive than least squares, it can be mitigated by multiple random initializations to avoid some negative local optimum.

2. The Robustness of Laplace EM

Laplace mixture regression changes this scenario in a couple of critical respects. First, each component of the mixture essentially employs an absolute-error loss. So, what this means is that an outlier in the Y direction doesn't matter how much its error value has; it just matters whether you are above or below that line. The attraction is only a constant push up or down, not a force that grows with distance, because the upper bound is +1 and the lower bound is -1. This automatically reduces its influence.

The Second, and even more powerfully, the mixture framework provides a layer of defense. For an outlier to impact a line of a given component, it itself has to be part of that component. For each point and each cluster, the model computes a posterior probability, which is a soft membership weight. A large outlier that doesn't fit the pattern of any cluster well will have very low membership weights for all components. This is essentially saying to the outlier, "you shouldn't be here," and potentially its ability to change the line of any component's line, exactly how much that is changed, or possibly even removed or neutralized. Let us focus on M-step, particularly on the weight function,

$$w_j^{(t)} = \frac{\sigma_i^{(t)}}{\sqrt{2}|y_j - x_j'\beta_i^{(t)}|} \qquad (6)$$

This function is a key idea of robustness against outliers because the denominator involves the absolute value of residuals. Surely, the presence of outliers makes large residuals, and the weight is very small, and vice versa is right. So, however large the outlier is, a Laplace Mixture model will have its in-sample amount at that fixed value. The contribution of the outlier is firstly reduced by the absolute error scheme and then eventually eliminated by the posterior membership probabilities. So, however large the outlier is, a Laplace mixture model will have its in-sample amount to that fixed value. The contribution of the outlier is firstly reduced by the absolute error scheme and then eventually eliminated by the posterior membership probabilities.

Unfortunately, the robustness of LAD can be violated in the presence of leverage points which can occur in the component as Type II, Type III and Type IV outliers. A simple definition of leverage point is an outlier in X-direction in regression model. A high-leverage point has a great deal of "pull" on the regression line. It

can totally obscure the slope. The high-leverage point lies on the line around the main center of the bulk of data, and the slope and intercept are way off. Although LAD is more resistant to outliers, a high-leverage point can still have a powerful effect and lead the estimates wrong. The LAD model attempts to minimize the vertical distance. If a point is far from zero on the X-axis, even a small angular error causes huge absolute residuals, and the model, therefore it is compelled to rotate towards that point. Consequently, the estimates are biased away from the truth. When the leverage point is present in the dataset, the desirable high breakdown point of LAD estimators drops to zero.

3. Weighted of EM for Laplace Mixture Regression

To overcome the problem of leverage point, our suggestion is described in the following steps,

1. Identify the leverage points using Robust Mahalanobis Distance (RMD) based on robust location $\hat{\mu}$ and scale and $\hat{\Sigma}$ of MCD or MVE.

$$RMD_i = \sqrt{(\mathbf{x}_i - \hat{\mu})^T \hat{\Sigma}^{-1} (\mathbf{x}_i - \hat{\mu})}, \quad i = 1,2,..,n$$

2. Computing Gilloni (2006) weights using

$$\zeta(\mathrm{x}_i) = \min\left(1, \frac{p}{RMD_i}\right)$$

3. Estimating EM algorithm parameters,

a) E-step

$$\tau_k^{(t)}(\mathrm{x}_i, y_i) = \frac{\pi_k^{(t)} \cdot \frac{1}{2\sigma_k^{(t)}\zeta(\mathrm{x}_i)} \exp\left(-\frac{|y_i - \mathrm{x}_i^T \beta_k^{(t)}|}{\sigma_k^{(t)}\zeta(\mathrm{x}_i)}\right)}{\sum_{j=1}^{K} \pi_j^{(t)} \cdot \frac{1}{2\sigma_j^{(t)}\zeta(\mathrm{x}_i)} \exp\left(-\frac{|y_i - \mathrm{x}_i^T \beta_j^{(t)}|}{\sigma_j^{(t)}\zeta(\mathrm{x}_i)}\right)}$$

b) M-step

$$\beta_k^{(t+1)} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \tau_k^{(t)}(\mathrm{x}_i, y_i) \cdot \zeta(\mathrm{x}_i) \cdot |y_i - \mathrm{x}_i^T \beta|$$

This is a weighted quantile regression at τ=0.5.

$$\sigma_k^{(t+1)} = \frac{\sum_{i=1}^{n} \tau_k^{(t)}(\mathrm{x}_i, y_i) \cdot \zeta(\mathrm{x}_i) \cdot |y_i - \mathrm{x}_i^T \beta_k^{(t+1)}|}{\sum_{i=1}^{n} \tau_k^{(t)}(\mathrm{x}_i, y_i)}$$

If the diagnostics method identifying all leverage points, $\zeta(\mathrm{x}_i)$. x_i is bounded, such that the upper bound is 1 and lower bound is then $\frac{p}{RMD_i}$, therefore the IF becomes bounded in both y and x directions. Finally, dual weights have to be present in the equation (5), the first one tackle the problem of leverage points and another weight in (6) to reduce the effect of outliers.

4. Simulation study

A detailed Monte Carlo study was carried out to examine the robustness properties of mixture regression estimators in the presence of leverage contamination. The generating mechanism is based on a two-component ($K = 2$), finite mixture of linear regression models under different distributional assumptions for comparison. We simulate $n = \{50,100,\}$independent observations for each Monte Carlo replication. The indicator $z_i \in \{1,2\}$ of a latent component follows multinomial distribution:

$$P(z_i = k) = \pi_k, \quad \text{where} \quad \boldsymbol{\pi} = (\pi_1, \pi_2) = (0.5,0.5), \quad k = 1,2.$$

Based on $z_i = k$, the response variable y_i is generated from a linear regression model:

$y_i = \mathbf{x}_i^{\top}\boldsymbol{\beta}_k + \varepsilon_{ik}$, where $\mathbf{x}_i = (1, x_{i1}, x_{i2})^{\top}$ is the vector of independent variables with an intercept term. These variables $\mathbf{x}_i$are generated identically and independently from standard normal distributions: $x_{i1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \quad x_{i2} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$.

The initial values of component-specific regression coefficients are,$\boldsymbol{\beta}_1 = (0,2,-1)^{\top}, \quad \boldsymbol{\beta}_2 = (1,-3,2)^{\top}$. The error terms ε_{ik} follow a symmetric Laplace distribution is generated from exponential distribution $e_i \sim \mathrm{Exp}(1/\sigma_k)$ times $b_i \sim \mathrm{Bernoulli}(0.5)$ such that $\varepsilon_{ik} = e_i . b_i$ with component-specific scales $\sigma_1 = 1$ and $\sigma_2 = 1.5$. Four types (I-TV) of I contamination are considered in this simulation with $\alpha = \{0.05,0.10,0.15,0.25\}$ the proportion of observations is contaminated by outliers and leverage points, equally split between the two components,

Let Ω denote as the set of contaminated indices, with $|\Omega| = n \times \alpha$. The Types contamination according to their affect can be generated as follows,

- Type I: $y_i^{\text{out}} = y_i + \eta_i, \quad \eta_i \sim \mathcal{N}(0, 10^2), \quad \forall i \in \Omega$.
- Type II: moderate response contamination generated $y_i^{\text{out}} = y_i + \xi_i, \quad \xi_i \sim \mathcal{N}(0, 4^2), \quad \forall i \in \Omega,$

  while the leverage points equally distributed for components,

for component 1, $x_{i1}^{\text{out}} = 20, \quad x_{i2}^{\text{out}} = x_{i2}, \quad \forall i \in), \quad \forall i \in \Omega \cap \{i: z_i = 1\},$

for component 2,$x_{i1}^{\text{out}} = x_{i1}, \quad x_{i2}^{\text{out}} = 20, \quad), \quad \forall i \in \Omega \cap \{i: z_i = 2\}.$

This simulation case would strongly influence both regression and covariance estimates.

- Type III: moderate leverage points with substantial y shifts

$x_{ij}^{\text{out}} = x_{ij} + \delta_{ij}, \quad \delta_{ij} \sim \mathcal{N}(10, 5^2), \quad \forall i \in \Omega,$ where $j = 1$ for $z_i = 1$ and $j = 2$ for $z_i = 2$.
$y_i^{\text{out}} = y_i + \zeta_i, \quad \zeta_i \sim \mathcal{N}(10, 12^2), \quad \forall i \in \Omega.$

- Type IV: The target of type is to generate extreme leverage points with small residuals compared to the incorrect model are produced, making them especially difficult for mixture estimation.

The target of this type is to create severe contamination by characterizing component misassignment with extreme leverage points covariates $\mathbf{x}_i^{\text{out}} = (1{,}10, -10)^{\top}, \quad \forall i \in \Omega$. The component label is intentionally flipped:

$$z_i^{\text{out}} = \begin{cases} 2 & \text{if } z_i = 1, \\ 1 & \text{if } z_i = 2, \end{cases} \quad \forall i \in \Omega.$$

The contaminated of y_i can be generated from the wrong component's model:

$$y_i^{\text{out}} = \mathbf{x}_i^{\text{out}\top} \boldsymbol{\beta}_{z_i^{\text{out}}} + \varepsilon_i^{\text{out}}, \quad \forall i \in \Omega,$$

where $\varepsilon_i^{\text{out}} = b_i \cdot e_i, \quad e_i \sim \text{Exp}\left(\frac{1}{\sigma_{z_i^{\text{out}}}}\right), \quad S_i = 2b_i - 1, \quad b_i \sim \text{Bernoulli}(0.5).$

The simulation experiment was carried out with $R = 5000$ independent Monte Carlo replications. For each replication $r = 1, \ldots, R$:

a) Generate a complete clean dataset $\mathcal{D}^{(r)} = \{(\mathbf{x}_i^{(r)}, y_i^{(r)})\}_{i=1}^{n}$.

b) Randomly select a subset $\Omega^{(r)}$ of size $|\Omega|$ and apply the contamination Type (I-IV).

c) Fit four competing mixture regression models to $\mathcal{D}^{(r)}$: Gaussian mixture regression (MixGau), Laplace mixture regression (MixLap), Weighted Laplace mixture regression (W.MixLap), and Student's t-mixture regression (Mixt)

d) Extract parameter estimates $\widehat{\boldsymbol{\theta}}^{(r)} = (\widehat{\boldsymbol{\beta}}_1^{(r)}, \widehat{\boldsymbol{\beta}}_2^{(r)}, \hat{\sigma}_1^{(r)}, \hat{\sigma}_2^{(r)}, \hat{\pi}_1^{(r)})$.

and then for each parameter $\theta_j \in \boldsymbol{\theta}$, finding the empirical performance metrics across replications:

$$\text{Bias}(\hat{\beta}_j) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\beta}_j^{(r)} - \beta_j^{\text{true}}).$$

$$\text{Bias}(\hat{\pi}_j) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\pi}_j^{(r)} - \pi_j^{\text{true}}).$$

$$\text{MSE}(\hat{\theta}_j) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\beta}_j^{(r)} - \beta_j^{\text{true}})^2.$$

$$\text{MSE}(\hat{\theta}_j) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\pi}_j^{(r)} - \pi_j^{\text{true}})^2.$$

Table 1 reports the MSE and bias of the parameter estimates when the sample size is 50 and the quantile contamination level is no more than 5%. This situation is challenging as well, due to the small sample size, which increases sensitivity both to misspecification of the model and outliers. The normal-mixture MLE does very badly under Type I outliers, which do not give leverage points on the response. The reported MSEs are orders of magnitude larger with significant bias, indicating that Gaussian likelihood-based EM estimation is entirely unrobust. Even a few response outliers can be enough to bias estimates of component-specific regression coefficients and mixing proportions. The Mixt method, on the other hand, which replaces the Gaussian component densities by heavy-tailed t-distributions, leads to a strong decrease in MSE as well as bias. Theoretically, we would expect this because the t-distribution will automatically downweight large residuals via its tail behavior. The MixLap method is also an improvement w.r.t MLE, which validates that Laplace errors are robust to response contamination, but its performance in this table in comparison to Mixt can be slightly better, suggesting that in the case where we have a very small sample size, the added flexibility of the t-distribution could be favorable at a low sample size. The W.MixLap does not dominate in case of the Type I outliers, where there is a lack of leverage contamination, and weights can bring less produce.

Table 1: MSE(Bias) of points estimate for all methods where $n = 50$ nd 5% outliers

N=50 , 0. 05	**MLE**	**Mixt**	**MixLap**	**W.MixLap**
Type I outliers				
β_{01}	35.0957 (0.999)	0.1606 (0.0113)	0.2595 (0.0474)	0.3143 (0.0424)
β_{11}	15.2629 (1.391)	0.691 (0.2356)	2.5297 (0.6585)	2.1424 (0.5506)

β_{21}	49.446 (1.4991)	0.5451 (0.1306)	1.4236 (0.4812)	1.2472 (0.3446)
β_{02}	0.3754 (0.1184)	0.1471 (0.0604)	0.5210 (0.1921)	0.2439 (0.1216)
β_{12}	3.9001 (0.7173)	0.3848 (0.0131)	1.7225 (0.3660)	1.6070 (0.3268)
β_{22}	2.0888 (0.5515)	0.4994 (0.0967)	1.0677 (0.2609)	0.9502 (0.1806)
π	0.0179 (0.0144)	0.0091 (0.0180)	0.0087 (0.0162)	0.0072 (0.0037)
Type II outliers				
β_{01}	1.2569 (0.3368)	1.1912 (0.5220)	0.8564 (0.3748)	0.1646 (0.0006)
β_{11}	6.0130 (2.1376)	4.4511 (1.7377)	5.2410 (2.1537)	1.8719 (0.5286)
β_{21}	2.2303 (1.2543)	1.4872 (0.9716)	1.4247 (1.1151)	0.7201 (0.1494)
β_{02}	1.3144 (0.5518)	3.8087 (1.0883)	1.3628 (0.4480)	0.2849 (0.1118)
β_{12}	5.0646 (1.3684)	5.4058 (1.6926)	6.9657 (2.3346)	1.8252 (0.5248)
β_{22}	2.4712 (0.7573)	3.0566 (1.1604)	3.6097 (1.5637)	1.0612 (0.3259)
π	0.0370 (0.0731)	0.0459 (0.0374)	0.0200 (0.0401)	0.0099 (0.0154)
Type III outliers				
β_{01}	1.0197 (0.2465)	0.3507 (0.1151)	0.3782 (0.0796)	0.1899 (0.0260)
β_{11}	3.7075 (0.9759)	2.6357 (0.8973)	3.7207 (1.1413)	3.0438 (0.5814)
β_{21}	1.8812 (0.6608)	1.7964 (0.6764)	1.3024 (0.7012)	2.0517 (0.5188)
β_{02}	0.3750 (0.2048)	0.9731 (0.3379)	0.9855 (0.3155)	0.2689 (0.0865)
β_{12}	3.8392 (0.8774)	3.7628 (0.8210)	4.9408 (1.2121)	2.4839 (0.6076)
β_{22}	0.9614 (0.2923)	1.3617 (0.4298)	1.7555 (0.6385)	1.4742 (0.1551)
π	0.0216 (0.0261)	0.0214 (0.0446)	0.0107 (0.0255)	0.0128 (0.0205)
Type IV outliers				
β_{01}	0.9279 (0.3233)	1.4297 (0.6583)	0.7291 (0.4653)	0.2718 (0.0241)
β_{11}	4.4289 (1.7607)	4.5988 (1.6999)	3.5818 (1.7435)	1.8849 (0.6714)
β_{21}	2.7226 (1.4248)	2.7325 (1.2817)	1.7927 (1.1285)	0.8937 (0.4665)
β_{02}	2.6915 (0.8008)	3.8976 (1.1073)	1.7981 (0.5855)	0.3525 (0.1778)
β_{12}	5.4840 (1.2879)	6.0742 (2.0753)	7.9131 (2.5614)	1.2609 (0.3842)
β_{22}	2.9536 (0.9585)	4.1078 (1.6148)	4.7602 (1.9368)	1.3077 (0.4699)
π	0.0355 (0.0871)	0.0412 (0.0387)	0.0278 (0.0070)	0.0087 (0.0245)

The scenario is reversed for Type II outliers when response and leverage are merged. In here it is where the weaknesses of Mixt and MixLap come out. While both are still robust to long-range dependence residuals, the influence of high-leverage observations can not be controlled, which results in higher MSE and volatile patterns for bias vs. parameter plots. On the other hand, W.MixLap always has the lowest MSE and bias,

showing that weighting based on leverage are necessary when contamination occurs in the design space. The situation behaves quite differently for Type III outliers (moderate leverage points), where the variances are considerably higher with all methods. Mixt and MixLap are not robust to parameter choice, trading superior performance based on the choice of parameters. However, W.MixLap keeps relatively lower MSE and less bias spread, indicating the stability might be better in the case that leverage is not too high. Contrary to the previous situations, for Type IV outliers (the worst contamination), MLE fails, Mixt and MixLap degenerate rapidly or are already vanished in our situation study, and W.MixLap dominates is not a direct inference; it confirms very clearly that this method has the advantage of some robustness in small samples with leverage contaminations.

Table 2: MSE(Bias) of points estimate for all methods where $n = 100$ and 5% outliers

N=100 , 0. 05	**MLE**	**Mixt**	**MixLap**	**W.MixLap**
Type I outliers				
β_{01}	0.1411 (0.0090)	0.0428 (0.0169)	0.0340 (0.0095)	0.0310 (0.0107)
β_{11}	3.2177 (0.6637)	0.7037 (0.1547)	0.0414 (0.0262)	0.0397 (0.0267)
β_{21}	1.2888 (0.4717)	0.1183 (0.0472)	0.0541 (0.0154)	0.0538 (0.0317)
β_{02}	0.2685 (0.1004)	0.0594 (0.0262)	0.0818 (0.0104)	0.0873 (0.0308)
β_{12}	3.7406 (0.6946)	0.5837 (0.0786)	0.0949 (0.0484)	0.1090 (0.0498)
β_{22}	1.4987 (0.5365)	0.2772 (0.0463)	0.1196 (0.0020)	0.1219 (0.0184)
π	0.0074 (0.0023)	0.0066 (0.0094)	0.0049 (0.0058)	0.0047 (0.0006)
Type II outliers				
β_{01}	1.1164 (0.3911)	2.7080 (1.2009)	1.5130 (0.7367)	0.0471 (0.0100)
β_{11}	5.6125 (2.0629)	3.7218 (1.8077)	4.4789 (2.0576)	0.0502 (0.0871)
β_{21}	1.8045 (1.1585)	1.5716 (1.0556)	1.5663 (1.1473)	0.0570 (0.0364)
β_{02}	1.7230 (0.6621)	2.6993 (1.1417)	1.5527 (0.4614)	0.0649 (0.0792)
β_{12}	4.7386 (1.6412)	6.5357 (2.2992)	8.4948 (2.8888)	0.4454 (0.1586)
β_{22}	2.7828 (0.9450)	3.3753 (1.6480)	3.7957 (1.9245)	0.3125 (0.1144)
π	0.0459 (0.0457)	0.0397 (0.0206)	0.0215 (0.0195)	0.0037 (0.0019)
Type III outliers				
β_{01}	0.1681 (0.0220)	0.2753 (0.2006)	0.1999 (0.1035)	0.0519 (0.0235)
β_{11}	3.7409 (1.1449)	1.4019 (0.6393)	1.8275 (0.7576)	0.0630 (0.0007)
β_{21}	2.2942 (0.9574)	0.8416 (0.5016)	1.1706 (0.6182)	0.0708 (0.0343)

β_{02}	1.6327 (0.5936)	0.7819 (0.2974)	0.2201 (0.2138)	0.0671 (0.0719)
β_{12}	5.1964 (1.3305)	3.4846 (1.0907)	3.5762 (1.0842)	0.1144 (0.0243)
β_{22}	1.9853 (0.5033)	1.3305 (0.7435)	1.3145 (0.7578)	0.1834 (0.1472)
π	0.0379 (0.0763)	0.0136 (0.0156)	0.0172 (0.0265)	0.0049 (0.0120)
Type IV outliers				
β_{01}	0.7441 (0.3621)	1.4215 (0.8034)	0.9297 (0.5762)	0.1303 (0.0739)
β_{11}	2.5941 (1.4600)	3.1971 (1.5601)	3.7240 (1.7742)	0.1364 (0.1803)
β_{21}	1.9771 (1.2800)	1.7407 (1.1298)	1.5002 (1.0683)	0.1161 (0.1704)
β_{02}	1.5659 (0.5383)	4.5189 (1.5243)	1.1891 (0.5114)	0.0928 (0.0708)
β_{12}	4.1425 (1.3151)	6.4841 (2.3221)	8.1674 (2.7524)	0.2613 (0.1198)
β_{22}	3.1407 (1.1865)	4.9476 (1.9824)	4.3831 (1.9643)	0.2817 (0.1748)
π	0.0346 (0.0580)	0.0338 (0.0857)	0.0085 (0.0008)	0.0068 (0.0134)

Table 2 considers the identical contamination pattern of Table 1 but with a small sample size (N = 100). A larger sample size leads to a considerable increase in estimation accuracy for all robust methods, but differences are still significant. Type I outliers: When the response contaminations are pure, Mixt, MixLap, and W.MixLap produce very small MSEs with little bias, meaning correctness under pure response contamination. Nevertheless, we observe that MLE has substantially larger MSE, indicating that the lack of robustness can't all be compensated for by increasing sample size. In Type II outliers, the advantages of W.MixLap are even more obvious. Under all settings, W.MixLap provides much smaller MSE values compared with any other method. Mixt and MixLap still suffer from leverage points sensitivity, and their biases are not minimal. This demonstrates that a heavy-tail approach is insufficient in the presence of leverage points. For the Type III outliers, Mixt results in a little bit better estimation than MixLap under some parameter choices, indicating that the mixing is useful for moderate leverage and heavy tails. However, W.MixLap always gives the most stable estimates with the least bias spread. With Type IV outliers, the conclusion is similar as table 1 but numerically improved: (W.MixLap) dominates, and Mixt and MixLap are weakened; again, MLE results are unusable. These results indicate that weighting is better as the sample size increases.

5. Conclusion

Such extensive simulation studies show that the robustness of the mixture regression model must address the residual outlier and the leverage point. This normal-mixture maximum likelihood estimate of the classical method alternative is very sensitive to various types of contamination and can perform even worse than

discarding low-level outliers. The techniques Mixt and MixLap benefit from their handling of the responses, respectively, thus reinforcing past research into heavy tails. However, this property of rendering them sensitive to leverage points makes these models impractical. In all cases, whatever sample sizes we choose or kind of outlier considered, W. MixLap's bias and MSE are less than those of its competitors' (except for a few occasions with Type III, in which the other two estimators perform better). Moreover, the W.MixLap method is not merely outperformed with 50 samples, but its superiority also holds in the case of 100 samples. That means the dual-weighted robustness is not only a question of finite samples but also an important requirement for practical mixture regression. Therefore, it can be considered as a method for dealing with contaminated data, and its approach is theoretically and empirically sound and well founded in application.

6. Recommendation

This work contributes to the theory of robust mixture modeling by establishing a formalism for encompassing high-breakdown point covariance estimation (MCD) within the EM algorithm's framework. This construction yields a new tool with provably improved robustness. The proposed procedure offers a computationally inexpensive and easy to use algorithm that retains the property of the Laplace distribution as being robust against heavy tailed noise, while resulting in substantially more reliable mixture regression fits obtained with contaminated data. The proposed algorithmic approach and the analysis developed provide a framework to design robust EM-like algorithms for other mixture families and error distributions.

Reference

1. Aitkin, M., & Wilson, G. T., Mixture models, outliers, and the EM algorithm. *Technometrics*, (1980). *22*(3), 325-331.
2. Bai, X., Yao, W., & Boyer, J. E. Robust fitting of mixture regression models. *Computational Statistics & Data Analysis*, (2012). *56*(7), 2347-2359.
3. Belsley, D. A. On the efficient computation of the nonlinear full-information maximum-likelihood estimator. *Journal of Econometrics*, *14*(2), (1980). 203-225.
4. Dempster, A. P., Laird, N. M., & Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society: series B (methodological)*, *39*(1), (1977). 1-22.
5. García-Escudero, L. A., & Mayo-Iscar, A. Robust clustering based on trimming. *Wiley Interdisciplinary Reviews: Computational Statistics*, *16*(4), (2024).e1658.
6. García-Escudero, L. A., Gordaliza, A., Matrán, C., & Mayo-Iscar, A. A general trimming approach to robust cluster analysis. (2008).

7. García-Escudero, L. A., Hennig, C., Mayo-Iscar, A., Morelli, G., & Riani, M. Choice of trimming proportion and number of clusters in robust clustering based on trimming. *Journal of Computational and Graphical Statistics*, (2025). ,1-13.
8. Hassan S. U. Weighted mixture regression model based on t-distribution in the presence of a leverage point. Mathematical Modeling and Computing. Vol. 12, No. 3, (2025), pp. 950–956
9. Hennig, C. ,"Fixed point clusters for linear regression: computation and comparison", Journal of classification, Vol. 19 , No. 2, (2002), pp. 249-276.
10. Hennig, C. Breakdown points for maximum likelihood estimators of locationscale, mixtures. *Annals of Statistics*, *32*, (2004). 1313-1340.
11. Lin, T. I., Lee, J. C., & Hsieh, W. J. Robust mixture modeling using the skew t distribution. *Statistics and computing*, *17*(2), (2007). 81-92.
12. McLachlan, G. J., & Peel, D. *Finite mixture models*. John Wiley & Sons (2000).
13. Quandt, R. E. A new approach to estimating switching regression. Journal of the *American Statistical Association*, *67*, (1972). 306-310.
14. Rousseeuw, P. J., & Leroy, A. M. *Robust regression and outlier detection*. John wiley & sons (2003). .
15. Song, W., Yao, W., & Xing, Y. Robust mixture regression model fitting by Laplace distribution. *Computational Statistics & Data Analysis*, *71*, (2014). 128-137.
16. Uraibi, H. S., & Waheed, M. Q. Weighted fast-trimmed likelihood estimator for mixture regression models. *Frontiers in Applied Mathematics and Statistics*, *10*, (2024). 1471222.
17. Yao, W., Wei, Y., &Yu, C. Robust mixture regression using the t-distribution. *Computational Statistics and Data Analysis*, 71, (2014). 116–127
18. Yu, C., Chen, K., & Yao, W. Outlier detection and robust mixture modelling using nonconvex penalized likelihood. *Journal of Statistical Planning and Inference*, (2015). 164, 27–38.